Children's social referencing reflects sensitivity to uncertainty

Emily Hembacher, Benjamin deMayo, Michael C. Frank

Department of Psychology, Stanford University

Author Note

Abstract

This study examined a spontaneous information gathering behavior – social referencing – and its relation to epistemic uncertainty during early childhood. Children ages 2-5 (n=160) identified the referents of familiar and novel labels. Referential ambiguity was manipulated through the number of objects present and their familiarity (Experiments 1 and 2), and the availability of referential gaze (Experiment 2). In both experiments, children visually referenced the experimenter more often while responding when the referent was ambiguous. In Experiment 2, 3- to 4-year-old children also referenced more when there was a novel referent and familiar distractor (and the referent could thus be inferred), but only when referential gaze was unavailable. These results suggest that children are able to seek disambiguating social information on the basis of graded uncertainty.

Keywords: social referencing; help seeking; word learning; uncertainty.

Children's social referencing reflects sensitivity to uncertainty

Preschoolers quickly learn new concepts, rules, and language. They also actively explore (Schulz & Bonawitz, 2007) and ask questions (Chouinard, Harris, & Maratsos, 2007) in ways that seem targeted to maximize learning. These behaviors appear to require children to monitor epistemic states of ignorance and uncertainty, but young children have generally been credited with limited ability to monitor mental states (Sodian, Thoermer, Kristen, & Perst, 2012). This conflict presents a puzzle – how do children acquire disambiguating information if they are not able to monitor uncertainty? To better understand how young children actively construct meaning from their environment, more evidence is needed about the relation between epistemic uncertainty and information seeking in early childhood.

Early research on metacognitive development suggested that young children have limited capacities to reflect on their own thoughts or knowledge. For example, 5-year-olds were found to have limited awareness of their own ongoing thoughts (J. H. Flavell, Green, Flavell, Harris, & Astington, 1995), and preschoolers were unable to recognize that they did not know all of the steps to complete a task unless prompted to act it out (Markman, 1977). Later studies focused on young children's ability to predict their performance in various tasks, often memory tasks, and found that children tended to be overly optimistic about their future performance (Schneider, 1998). These and other studies tended to rely on children's open-ended verbal reports about their mental operations, however, and such measures might underestimate children's ability to track epistemic states.

More recently, researchers have investigated children's uncertainty monitoring using tasks that do not require a verbal response from children. In one study, 3- to 5-year-olds completed perceptual discrimination and lexical identification tasks and

reported on their certainty in their choices using a pictorial confidence scale (Lyons & Ghetti, 2011). Preschoolers reported being less confident when they responded incorrectly, suggesting that they were aware of their likelihood of accuracy based on their epistemic states. In another study, 3.5-year-olds completed a memory task in which they could opt out of responding to individual trials (Balcomb & Gerken, 2008). Children performed worse on trials they had opted out of when they were forced to answer them later on, suggesting that they used the opt-out option strategically to avoid answering when they were uncertain about their responses.

On the other hand, in a study with a similar structure that investigated children's confidence in their memories for pictures, 3-year-olds were equally confident for correct and incorrect answers, while 4- and 5-year-olds were more confident for their correct answers (Hembacher & Ghetti, 2014). Furthermore, only 5-year-olds were most confident about their memories for images they had studied for longer. These results suggest that there are limits to 3-year-olds' abilities to monitor uncertainty, and that some cognitive states may be more easily reflected on by young children than others. The developmental trend also suggests that there is improvement in children's uncertainty monitoring between ages 3 and 5.

In sum, preschool-aged children demonstrate an emerging but limited awareness of epistemic uncertainty in laboratory tasks that do not require a verbal or open-ended response. However, even the competencies reported in these studies may underestimate children's ability to track epistemic uncertainty and act on it. One possibility is that children's tracking of uncertainty is apparent in their spontaneous information-seeking behaviors before they can report on it explicitly. Such a possibility would mirror the pattern proposed in other domains, e.g., Theory of Mind (**???**).

In the following sections, we first outline an emerging body of research that

suggests that children actively gather data to support learning from a young age, in a way that reflects at least an implicit sensitivity to epistemic uncertainty. We then summarize previous work on children's ability to gather social information selectively from credible sources and under conditions of epistemic uncertainty.

Active Learning in Early Childhood

Active learning refers broadly to learning behaviors and materials that are initiated or generated by the learner themself, rather than by a teacher (Gureckis & Markant, 2012). Evidence for active learning during early childhood comes from studies of overt question-asking and help-seeking in preschoolers and toddlers, and behaviors that reflect active information-seeking in infants.

Spontaneous exploration in preschoolers and toddlers. First, a collection of studies in the field have taken inspiration from comparative metacognition research and asked whether children's spontaneous information-seeking behaviors track uncertainty. For example, Call and Carpenter (2001) had 2-year-olds choose between several tubes to find a hidden sticker. They found that the toddlers were more likely to peek inside a tube before choosing when they had not seen the baiting of the tubes compared to when they had, suggesting they were aware of their knowledge or ignorance and selectively sought confirmatory evidence before committing to a response when they did not know a sticker's location. Others have found similar patterns of results (**???**). In another study, Goupil, Romand-Monnier, and Kouider (2016) trained 20-month-olds to look at their parents if they needed help with a memory task in which they had to locate a hidden toy. Toddlers were more likely to seek help when they were completely ignorant of the toy's location, and when the task was more difficult because the delay between hiding and test was longer.

There is also evidence that young children selectively explore when evidence is confounded (i.e., when existing evidence supports multiple causal hypotheses). For example, Schulz & Bonawitz (2007) found that preschoolers spent more time playing with a toy when they witnessed confounded evidence about its causal structure. Children who experienced a demonstration of a toy in which two levers were simultaneously depressed and caused two simultaneous events (a toy duck and puppet popping up) spent more time playing with that toy compared to children who saw a demonstration in which one lever was depressed at a time, revealing the unique function of each lever. This finding is consistent with the prediction that preschool-aged children track causal ambiguity and spontaneously explore more when they do not have sufficient evidence to isolate a cause for an effect.

Young children are also capable of independently generating the evidence they need to distinguish between causal hypotheses. Sim and Xu (2017) found that 2- and 3-year-olds were able to spontaneously discover first- and second-order rules to activate a toy (e.g., the triangle block makes the triangle toy work and toys are activated by shape-matched blocks) during free-play. Preschoolers also identify the most informative questions to confirm or disconfirm hypotheses (Ruggeri, Sim, & Xu, 2017). In both cases, these behaviors would be unlikely to occur if children did not have some sensitivity to their epistemic states; children must thus track uncertainty in order to intervene effectively and gather disambiguating data.

Information-seeking and uncertainty-tracking in infants. Infants have also been shown to seek information selectively under conditions of ambiguity, and explore or allocate attention to information sources that are most likely disambiguate. For example, in one pair of studies, 7- and 8-month-olds attended longer to visual displays (Kidd, Piantadosi, & Aslin, 2012) or auditory stimuli (Kidd, Piantadosi, &

Aslin, 2014) that were neither too complex nor too simple, and thus afforded the greatest learning opportunity. Twelve-month-old infants have also been shown to track the likelihood of different events (e.g., a red ball exiting a container full of moving red and blue balls) based on reasoning about multiple cues – numerosity (the number of red compared to blue balls), physical distance (the distance from a red ball to the opening of the container) and time (how long the scene was occluded before the ball left the container). Intriguingly, infants' surprisal (looking time) tracked with uncertainty about the likelihood of events – for example, at intermediate occlusion times, infants' looking time was graded with respect to distance and numerosity in an additive fashion (Teglas et al., 2011).

While most studies of information-seeking in infancy have relied on looking behavior, Stahl and Feigenson (2015) explored infants' exploratory actions with toys that had either behaved normally or appeared to violate physics (e.g., by hovering in midair). They found that infants were more likely to explore those objects that had violated physical expectations, and they explored in ways that were suited to testing the physical property they had seen violated (e.g., picking up and dropping the car that had hovered in midair; banging a toy that had appeared to move through a solid wall). These results suggest that even young infants selectively explore objects to disambiguate causal properties that have been rendered uncertain by a surprising event.

In sum, there is a growing body of evidence that infants and young children are active learners who explore based on their epistemic states. These studies also suggest that spontaneous information-seeking behaviors may be a fruitful behavioral index of children's sensitivity to uncertainty in different learning contexts. However, much of this research is confined to situations in which children must discover the causal properties of interesting objects. Does epistemic uncertainty drive other types of information seeking

across learning domains and information sources?

In particular, we focus here on the role of uncertainty monitoring in children's social learning. Infants and children learn a great deal from social partners, through observation and direct pedagogy. But not all social information is created equal, and learners must determine *who* to seek social information from, *when* to seek it, and *how* to seek it. In the next section we outline previous research on the selectivity of early social learning.

Selective social learning

Although children appear to be biased towards trusting others' testimony overall (Jaswal, Croft, Setia, & Cole, 2010), there is evidence that they are selective in their social learning. Children track the quality of informants' testimony, and prefer to learn from people who have a history of accuracy (Corriveau & Harris, 2009; Sobel & Kushnir, 2013). For example, older preschoolers choose to learn the words for novel objects from someone who has previously labeled objects correctly compared to incorrectly (Koenig, Clement, & Harris, 2009), and from people who appear confident rather than uncertain (Sabbagh & Baldwin, 2001). They also prefer to learn from people who have proven to be honest rather than dishonest in the past (Li, Heyman, Xu, & Lee, 2014)]. Altogether, young children appear to be selective learners with regard to the *sources* of social information they seek.

Are children also selective about *when* they seek social information? That is, do they seek social information on the basis of their uncertainty about their own mental representations? There is some evidence to suggest that they do, at least in the context of solving problems of varying difficulty. For example, Vredenburgh & Kushnir (2015) found that preschool-aged children selectively sought help with a task when they were

either less competent at the task or when they reached a difficult step. Coughlin, Hembacher, Lyons and Ghetti (2014) found that preschoolers were more likely to seek help on trials of a perceptual discrimination task for which they reported lower confidence in a separate session in which help was not available. Additionally, an analysis of a corpus of parent-child conversations showed that preschoolers tend to ask questions that are directed to resolve gaps in knowledge, and they ask follow-up questions if they receive uninformative responses (Chouinard et al., 2007).

Overt help-seeking is only one form of social information-gathering. Infants and children also engage in *social referencing*, or checking a social partner's face for gaze direction and/or reactions to stimuli or events (Walden & Ogan, 1988). There is plentiful evidence that infants reference social partners to determine the safety of objects or actions; for example, infants reference their parents' faces before crossing a bridge or drop-off of uncertain slope or depth (Sorce, Emde, Campos, & Klinnert, 1985; Tamis LeMonda et al., 2008). Infants also check adults' faces when exposed to potentially dangerous stimuli, such as barking sounds or spiders (Striano & Rochat, 2000; Zarbatany & Lamb, 1985). In summary, infants and young children reference social information in response to uncertainty about the physical affordances or safety of a situation.

Social referencing is especially critical for language acquisition. By the second year of life infants follow a speaker's gaze and map labels to objects on the basis of gaze direction (Baldwin, 1991). There is also evidence that infants' propensity for gaze-following predicts later language development, highlighting the importance of this behavior for learning. Furthermore, word learning is more successful under joint attention, which involves mutual knowledge about a shared focus of attention (Baldwin, 1993; Tomasello & Farrar, 1986).[1]

———————

[1]We use the term "social referencing" to indicate the generic behavior of seeking social information

However, there is still only limited evidence about whether infants and young children are *selective* in seeking out these social cues to reference. Do children orient towards a speaker indiscriminately, or do they selectively look for social cues when reference cannot be determined without them? It could be that social referencing is typically not costly enough to require selectivity. People shift gaze several times per second, and social cues are intrinsically salient (Csibra & Gergely, 2006). Thus, children may sample social information frequently regardless of its potential to disambiguate reference.

Another possibility is that uncertainty signals related to knowledge representations are too weak to drive information-seeking behaviors in young children. Consistent with this possibility, as discussed previously, studies of early metacognition have typically found that young children are overconfident in their knowledge and performance predictions, even when they are calibrated overall to accuracy (Coughlin et al., 2014; Hembacher & Ghetti, 2014; Schneider, 1998). If children are not aware of the need for disambiguating information, they may not selectively reference social information based on referential ambiguity.

At least one study has previously investigated the selectivity of infants' social referencing during word learning. Infants heard an experimenter produce a label for an object in the presence of one or two novel objects (Vaish, Demir, & Baldwin, 2011). Infants looked up at the experimenter more often when there were two objects present, and the referent was thus ambiguous. While this study provides encouraging evidence that infants' social referencing may be sensitive to referential uncertainty, there are questions that remain unanswered. Since referential ambiguity was manipulated through the number of items present, it is unclear whether infants' additional looking was based

_____

whether it is specifically referential (as in gaze following) or otherwise.

on epistemic uncertainty or the complexity of the environment (i.e., the presence of multiple objects). In the current study, we seek to replicate these findings with preschool-aged children while controlling for the number of objects present.

Current study

The present study builds on previous work suggesting that preschoolers are active and selective in their social learning. Here, we ask whether preschoolers reference a speaker more frequently when the referent of their speech is ambiguous. To this end, we adapt the procedure used by Vaish et al. (2011) for use with young children, controling for the number of objects present. In Experiment 1 we include trials with two novel objects (referentially ambiguous), two familiar objects (referentially unambiguous), one novel object (referentially unambiguous) and one familiar object (referentially unambiguous). In Experiment 2, we additionally manipulate the amount of referential evidence available by manipulating whether or not the speaker's gaze is informative, allowing us to investigate whether preschoolers' social referencing is sensitive to graded uncertainty.

There are several reasons that investigating children's sensitivity to graded evidence is important. First, probabilistic models of cognition and learning require that individuals be able to track the amount of evidence for alternative beliefs (and corresponding uncertainty), perhaps through heuristics such as "win-stay lose-sample" (E. Bonawitz, Denison, Gopnik, & Griffiths, 2014). If preschool-aged children do track accumulating evidence as assumed by these models, their information-gathering behaviors should be graded with respect to the amount of evidence for a belief. Critically, although individuals may not adjust their information seeking in a graded fashion with respect to uncertainty, situations characterized by intermediate referential

uncertainty (e.g., because only one cue is present) should elicit intermediate amounts of information-seeking in the aggregate (Daw & Courville, 2007).

More generally, a complete theory of early metacognition and learning should encompass children's ability to respond appropriately to graded evidence. Metacognitive theories assume that effective behavioral regulation and response selection rely on individuals' abilities to track the likelihood of being correct based not only on complete knowledge or ignorance, but on graded evidence (Lyons & Zelazo, 2011). For example, having more information in support of a belief should make individuals more willing to share that information with others, and having less information should motivate help-seeking (Ghetti, Hembacher, & Coughlin, 2013).

Experiment 1

In Experiment 1, we examined whether children would visually reference a speaker more often when the speaker produced a referentially ambiguous label compared to an unambiguous label. Children sat across from an experimenter who labeled an object on the table between them (Figure 1). The experimenter then asked the child to place the named object in a bucket. Across trials, there were either one or two objects on the table, which were either familiar or novel to the child. This design allowed us to test whether merely having more than one object present is sufficient to increase social referencing, or if referential ambiguity (and thus epistemic uncertainty) is the underlying factor. If the latter is true, we expected children to increase their looking to the experimenter only on trials with two unfamiliar objects, when the object-label mapping was not known and could not be inferred.

We were interested in the amount of social referencing children exhibited across the trial. We considered four different phases of each trial based on the notion that

children might expect different social information at different stages of the task. Specifically, we predicted that children might expect the speaker's gaze direction to be informative during the labeling itself, as speakers tend to look at objects they refer to. We also predicted that later in the trial, as children reached for an object and placed it in the bucket, they might expect evaluative feedback about their choice (e.g., facial expressions of encouragement or discouragement).

Methods

Participants. We recruited a planned sample of 80 children ages 2-5 years from the Children's Discovery Museum in San Jose, California.[2] The sample included 20 2-year-olds (mean age 32.0 months), 20 3-year-olds (mean age 42.7 months), 20 4-year-olds (mean age 55.8 months), and 20 5-year-olds (mean age 65.2 months). An additional 20 children participated but were removed from analyses because they heard English less than 75% of the time at home ($n = 10$), because they were unable to complete at least half of the trials in the task ($n = 4$), because of parental interference ($n = 1$), or due to experimenter or technical errors ($n = 5$).

Stimuli and Design. Children were presented with one or two objects, heard a label, and were asked to put the labeled object in a bucket. Half of the objects were selected to be familiar to children (e.g., a cow) and half were selected to be novel (e.g., a nozzle). There were four possible trial types based on the number and familiarity of the objects present: one familiar object (F), one novel object (N), two familiar objects (FF), and two novel objects (NN). There were three trials of each type, for a total of twelve trials. Trial types were presented sequentially in an order that was counterbalanced across participants. The assignment of individual objects to trial types was

---

[2]Planned sample size, exclusion criteria, and analysis plan preregistered at `https://osf.io/y7mvt`.

counterbalanced. On F/FF trials, the familiar label for the target object was used (e.g., "cow"). On N/NN trials, a novel label was used (e.g., "dawnoo").

The critical manipulation was of referential ambiguity; F and FF trials were referentially unambiguous, as children were expected to be certain about the objects and their labels. Similarly, on N trials, children were expected to be certain about the label referent as there was only one option. However, NN trials were referentially ambiguous, as the novel label could apply to either novel object.

Throughout the task, the experimenter never gazed at the object they were labeling, or responded to children's verbal or non-verbal bids for help by indicating the correct object. Thus, children were expected to remain uncertain about the referent for the duration of the trial when two novel objects were present.

Procedure.   Throughout the study, the child sat at one end of a large circular table, and the experimenter stood at the opposite end. Each trial proceeded as follows: the experimenter placed one or two objects on the left and/or right sides of the table, out of reach of the child so that the child could not interact with the toys during the labeling event. For one-object trials, the location of the object (left or right) alternated between trials.

After placing the objects, the experimenter said "Hey look, there's a (target) here." The experimenter gazed at the center of the table rather than the object they labeled. The experimenter waited approximately two seconds based on a visual metronome placed within view before saying, "Can you put the (target) in the bucket?" They then pushed the object(s) forward within reach of the child, and placed a plastic bucket in the center of the table, also within reach of the child. Prior to the twelve experimental trials, there were two training trials: a F trial and a FF trial, to acquaint the child with the procedure. A camera placed to the side of the experimenter captured the

participant's face, so that looking behavior could be coded from video.

Coding procedure and analytic plan. Videos were coded using DataVyu software (`http://datavyu.org`). For each participant, we coded the *number of times* they referenced the experimenter throughout each trial. An alternative analytic option would be to simply code *whether or not* children looked at the experimenter. However, during piloting, we found that most children looked up to the experimenter at least once while they were labeling the object, suggesting that a binary measure of looking would not be meaningful.

Because we were interested in the precise circumstances in which children feel uncertain enough to reference a speaker, we coded the number of looks that occurred during four distinct phases of the trial: a *label* phase, which began at the utterance of the label and ended when the experimenter began to slide the objects, a *slide* phase, in which the experimenter slid the object(s) into the child's reach, a *planning* phase, which began at the end of the slide and ended when the child touched an object, and a *response* phase, which began when the child touched an object and ended when the child released the object into the bucket.

We also noted any trials that should be excluded from analyses due to the child's interference with the timing of the trial (e.g., talking over the experimenter), experimenter error, or outside distractions that interfered with the timing of the trial (e.g., noise from a sibling). These trial-wise exclusion criteria were preregistered. In total, we excluded 1.4% of trials from 2-year-olds, 2.0% of trials from 3-year-olds, 1.9% of trials from 4-year-olds, and 2.0% of trials from 5-year-olds on this basis.

A second coder independently scored the number of looks for one third of the trials for each participant to establish reliability. Inter-rater reliability for the number of looks in each phase was high, intraclass correlation $r = .97$, $p<.001$.

Table 1 displays the average durations of each of the four phases. The *label* and *response* phases are longer on average than the *planning* and *slide* phases. However, note that we are interested in comparing the amount of social referencing across ambiguity conditions and not across phases directly.

To quantify the effects of the number and familiarity of objects on children's looking, along with any developmental trends, we planned to fit a linear mixed-effects regression, beginning with the following structure and trimming according to standard laboratory procedures[3]: `number of looks ~ number of objects * familiarity * phase * age in months + (number of objects * familiarity | subject)`. This model specification was preregistered, as noted above.

Results and Discussion

[3]Standard laboratory analytic procedures available at `https://osf.io/zqzsu/wiki/home/`

	Experiment	Phase	Mean Duration	SD
1	1	label	5072.84	899.16
2	1	slide	872.43	196.83
3	1	planning	1253.85	2215.56
4	1	response	3646.40	4510.01
5	2	label	5239.23	748.05
6	2	slide	818.95	1307.27
7	2	planning	790.88	10327.45
8	2	response	4084.56	5032.37

Table 1

Phase Durations

Accuracy. We examined children's accuracy for those trials in which a correct response was possible (i.e., FF trials). Children sometimes put two items in the bucket (2-year-olds: 30.8% of 2-object trials; 3-year-olds: 15.0%; 4-year-olds: 7.5%; 5-year-olds: 9.2%), despite instructions to choose only one. Children's responses were coded as the first object that they placed in the bucket. Children in each age group were equally likely to put two objects in the bucket for familiar and unfamiliar trials (2-year-olds: $t(19) = 0.66$, $p = 0.52$; 3-year-olds: $t(19) = 1.45$, $p = 0.16$; 4-year-olds: $t(19) = -0.72$, $p = 0.48$; 5-year-olds: $t(19) = 0.37$, $p = 0.72$) suggesting that they put two objects in the bucket because it was a fun activity and difficult to inhibit, and not because they did not know the referent. Among FF trials in which children placed two items in the bucket, they put the correct item in first about half of the time (8/19 trials total for 2-year-olds; 5/6 for 3-year-olds; 0/2 for 4-year-olds; 2/4 for 5-year-olds).

Children also occasionally declined to choose an item (0.4% of trials); these trials are excluded from accuracy analyses. Children's accuracy is displayed in Figure 2. Overall children generally chose the correct item for FF trials, indicating that they understood the task and were motivated to answer correctly. While 3- to 5-year-olds performed close to ceiling (92% - 99%), 2-year-olds were less accurate (76%), but still performed significantly above chance ($t(19) = 23.0$, $p < .001$).

Social referencing. Results of Experiment 1 are presented in Figure 3. To test our prediction that referential ambiguity (i.e., having two novel objects) would produce more social referencing, we fit mixed-effects linear regression models separately for each phase with the following structure: `number of looks ~ number of objects * familiarity * age in months + (number of objects + familiarity | subject)`. A single model with phase as a factor did not converge, and the model was subsequently trimmed according to our standard laboratory analytic procedures.

We did not find any main or interaction effects of number of objects, familiarity, or age on number of looks during the *label* or *slide* phases. Thus, mere novelty or the presence of multiple objects was not enough to increase social referencing. However, we found an interaction effect of number of objects and familiarity during the *planning* (β = 0.2, $p < .001$) and *response* phases ($\beta = 0.6$, $p < .001$), such that NN trials were associated with more looking.

There was no interaction with age in any phase.[4] Since an age trend in the *planning* phase was apparent in the plot (Figure 3), we conducted a post-hoc mixed-effects linear regression with age group (2- and 3-year-olds vs. 4- and 5-year-olds) predicting the number of looks in the *planning* phase: `number of looks ~ number of objects * familiarity * age group + (number of objects + familiarity |`

[4]`https://github.com/emilyfae/socref_uncert`

Predictor	Estimate	Std. Error	t value	p value	
Intercept	1.38	0.06	22.58	< .001	***
Num objects (2)	-0.06	0.06	-1.02	0.31	
Familiarity (N)	0.07	0.06	1.16	0.25	
Age	0.01	0.00	1.85	0.07	.
Num objects (2) * Familiarity (N)	-0.06	0.09	-0.74	0.46	
Num objects (2) * Age	-0.01	0.00	-1.18	0.24	
Familiarity (N) * Age	-0.01	0.00	-1.17	0.24	
Num objects (2) * Familiarity (N) * Age	-0.00	0.01	-0.57	0.57	

Table 2

Predictor estimates with standard errors and significance information for a linear mixed-effects model predicting social referencing in the label phase in Experiment 1.

`subject)`. We found a significant interaction of number of objects and familiarity ($\beta =$ 0.3, $p < .001$), and a significant 3-way interaction between number of objects, familiarity, and age group, such that younger children looked at the speaker less than older children for NN trials. Although this was a post-hoc analysis, it suggests that there may be a trend for children to respond more quickly to uncertainty as they grow older. This could mean that they detect uncertainty more quickly, or that they are more proactive in seeking disambiguating information prior to initiating a response.

As discussed previously, an alternative analytic approach would be to fit a logistic mixed-effects regression predicting *whether or not* children look to the speaker in each phase, rather than the number of times they do so. In support of this approach, the distribution of looks is not normal for the slide, planning, and response phases, with the majority of trials containing no looks to the speaker (Figure 4). To address this issue,

Predictor	Estimate	Std. Error	t value	p value	
Intercept	0.12	0.02	5.06	$< .001$	***
Num objects (2)	0.02	0.03	0.55	0.58	
Familiarity (N)	0.04	0.03	1.38	0.17	
Age	-0.00	0.00	-0.44	0.66	
Num objects (2) * Familiarity (N)	-0.07	0.04	-1.69	0.09	.
Num objects (2) * Age	0.00	0.00	0.20	0.84	
Familiarity (N) * Age	-0.00	0.00	-0.18	0.86	
Num objects (2) * Familiarity (N) * Age	-0.00	0.00	-0.84	0.4	

Table 3

Predictor estimates with standard errors and significance information for a linear mixed-effects model predicting social referencing in the slide phase in Experiment 1.

we additionally fit separate logistic mixed-effects regressions with the following structure: `look (yes or no) ~ number of objects * familiarity * age group + (number of objects + familiarity | subject)`. A single model with phase as a factor did not converge.

The results of these analyses were similar to those of the linear models; number of objects and familiarity interacted in the *planning* and *response* phases such that there was more likely to be a look when there were two novel objects, (βs $= 1.4$, 1.9, ps $<.05$ & $<.001$). In addition, there was a significant interaction of number and familiarity of objects in the *slide* phase, such that NN trials were less likely to be associated with a look ($\beta = -0.9$, $p <.05$). This latter finding was not predicted, but could reflect children's tendency to look more at novel stimuli, which would trade off with looking at the speaker, particularly when there is no utility to looking at the speaker, as when the

Predictor	Estimate	Std. Error	t value	p value	
Intercept	0.03	0.02	1.48	0.14	
Num objects (2)	0.01	0.04	0.26	0.8	
Familiarity (N)	0.00	0.03	0.12	0.9	
Age	-0.00	0.00	-0.92	0.36	
Num objects (2) * Familiarity (N)	0.20	0.05	4.34	< .001	***
Num objects (2) * Age	0.00	0.00	0.54	0.59	
Familiarity (N) * Age	0.00	0.00	0.73	0.47	
Num objects (2) * Familiarity (N) * Age	0.00	0.00	1.42	0.16	

Table 4

Predictor estimates with standard errors and significance information for a linear mixed-effects model predicting social referencing in the planning phase in Experiment 1.

objects are being slid across the table.

In summary, children looked to the speaker more often when planning and executing a response under uncertainty. These results suggest that children were aware that they did not have sufficient knowledge to answer independently, and referenced the speaker to resolve this uncertainty.

Notably, and in contrast to Vaish et al., we did not find the expected effect of referential ambiguity in the *label* phase. It is possible that children failed to predict that they would need more information until later in the trial, when they were actually faced with making a decision. Another possibility is that children's looking was at ceiling during the labeling phase, perhaps because children tend to look at someone who is speaking regardless of the need for referential disambiguation.

A third possibility is that this finding is an artifact of our design, in which the

Predictor	Estimate	Std. Error	t value	p value	
Intercept	0.22	0.04	5.35	< .001	***
Num objects (2)	0.05	0.06	0.96	0.34	
Familiarity (N)	0.01	0.05	0.18	0.86	
Age	-0.00	0.00	-0.20	0.84	
Num objects (2) * Familiarity (N)	0.60	0.07	8.21	< .001	***
Num objects (2) * Age	0.00	0.00	0.85	0.4	
Familiarity (N) * Age	0.00	0.00	0.55	0.58	
Num objects (2) * Familiarity (N) * Age	0.01	0.01	1.34	0.18	

Table 5

Predictor estimates with standard errors and significance information for a linear mixed-effects model predicting social referencing in the response phase in Experiment 1.

speaker gazed at the center of the table rather than the referent of the label. Children may have realized that the speaker's gaze direction during labeling was not informative. Alternatively, children may have found it strange to interact with a speaker who did not gaze at the object they were labeling, which may have produced unnatural patterns of social referencing. Experiment 2 tests these possibilities and examines whether children's social referencing is sensitive to graded uncertainty.

Experiment 2

Experiment 2 was designed to achieve several goals; first, we aimed to replicate the finding from Experiment 1 that children reference a social partner on the basis of referential ambiguity while executing a decision. Second, we tested whether children's social referencing is graded with respect to graded evidence about a label's referent.

To this end, we manipulated two sources of referential evidence. First, we added trials with 1 novel and 1 familiar object (FN) and a novel label. This condition contains evidence about reference since the familiar item can be excluded (Markman & Wachtel, 1988), but may not be as conclusive as trials with a familiar target. Thus, we predicted that, in the aggregate, children would show the most looking to the speaker in the NN condition, the least in the FF condition, and an intermediate amount in the FN condition.

Second, we manipulated between participants whether or not the speaker gazed at the objects they referred to, and thus, whether or not their gaze was an informative cue to reference. We predicted that having access to referential gaze as an informative cue would make children less likely to reference the speaker during their decision, but perhaps more likely to reference the speaker during labeling. Critically, this also allowed us to test whether children's social referencing is selective on the basis of referential

ambiguity during labeling if the speaker's gaze is informative, addressing an interpretive issue in Experiment 1.

Since we did not observe any difference between F and N trials in Experiment 1, we eliminated single-object trials, leaving the FF and NN trials. Additionally, we restricted the sample to 3- and 4-year-olds, as there was no effect of age in our preregistered analyses in Experiment 1, but a trend for younger children to engage in selective referencing earlier in the trial than older children.

Methods

Participants. We recruited a planned sample of 80 children ages 3-4 years from the Children's Discovery Museum in San Jose, California.[5] The sample included 40 3-year-olds (mean age 42.9 months) and 40 4-year-olds (mean age 53.5 months). An additional 20 children participated but were removed from analyses because they heard English less than 75% of the time at home ($n = 9$), because they were unable to complete at least half of the trials in the task ($n = 7$), or due to experimenter or technical errors ($n = 4$).

Stimuli and Design. The stimuli and design were similar to Experiment 1 but included three trial types: FF, NN, and FN. There were four of each trial type, totaling twelve trials. In addition, we manipulated the experimenter's gaze behavior between participants. For half of the participants, the experimenter gazed at the center of the table while labeling objects (uninformative gaze); for the remaining half, they gazed directly at the objects they labeled (informative gaze).

---

[5]Planned sample size, exclusion criteria, and analysis plan (including model specification) preregistered at `https://osf.io/y7mvt/`.

Procedure.   The experimental and coding procedures were identical to Experiment 1, except that there were three practice trials (two familiar trials and one novel trial). We chose this approach so that children could experience both familiar and novel stimuli during practice, but would not be discouraged by an overly difficult practice session.

We again noted trials that should be excluded based on the same criteria as in Experiment 1. We excluded 1.9% of trials from 3-year-olds and no trials from 4-year-olds. Inter-rater reliability for the number of looks in each phase was again high, intraclass correlation $r = .97$, $p<.001$.

The mean durations of the phases for Experiment 2 are presented in Table 1. They varied in length according to the same pattern as in Experiment 1.

Results and Discussion

Accuracy.   Children's accuracy for trials with a correct answer (i.e., FF, FN, and all trials with gaze) was calculated using the same criteria as in Experiment 1 (Figure 5). As in Experiment 1, children sometimes put two items in the bucket (3-year-olds: 0.0% of trials; 4-year-olds: 0.0%). If they put the correct object in first, the trial was counted as accurate. To quantify the effects of familiarity, gaze, and age on accuracy, we fit the following mixed-effects logistic regression model: `correct ~ trial type * gaze + age in months + (1| subject)`. Accuracy was significantly lower for novel (β = -2.3, $p < .001$) and mutual trials (β = -2.2, $p < .001$) compared to familiar trials, and trial type interacted with gaze condition such that accuracy was significantly higher in the gaze condition for novel ($\beta = 3.4$, $p < .001$) and mutual trials ($\beta = 1.1$, $p < .001$). Age did not significantly predict accuracy ($\beta = 0.2$, $p = .12$).

Social referencing and referential ambiguity. Children's social referencing based on trial type and gaze condition are presented in Figure 6. To quantify the main and interactive effects of familiarity, gaze informativity, phase, and age on social referencing, we fit a mixed-effects linear regression model with the following structure: `number of looks ~ familiarity * age in months * gaze * phase + (familiarity | subject)`. In contrast to Experiment 1, a model with phase as a predictor converged.

First, do children reference a speaker more often when the objects and label are novel? Phase interacted with familiarity such that the *response* phase of NN trials was associated with significantly more looks ($\beta = 0.3$, $p < .001$). This result is consistent with results from Experiment 1. However, in contrast to Experiment 1, the number of looks was not significantly greater for NN trials in the *planning* phase.

We were also interested in whether FN trials would elicit an intermediate amount of looking. We observed a three-way interaction of familiarity, gaze, and phase, such that the response phase of FN trials in the no-referential-gaze condition was associated with significantly more looks ($\beta = $ -0.2, $p < .01$). Thus, FN trials were associated with greater looking only when the experimenter did not provide informative gaze. This finding is intriguing given that children should be able to solve FN trials without gaze information. Instead, these results suggest that children remain relatively uncertain while making a decision if excluding the familiar object is their only cue to reference, but this uncertainty is resolved if the speaker's gaze is informative. On the other hand, informative gaze during labeling did not lessen social referencing for novel trials, suggesting that gaze information alone was not sufficient to reduce uncertainty. Instead, children required both types of evidence to feel certain about their response.

Finally, we observed a four-way interaction such that the *response* phase of novel

trials in the gaze condition was associated with more looking with increasing age (β = 0.0, $p < .01$), suggesting that preschoolers may become more selective in their social referencing as they get older. It may be that children improve in their ability to monitor the need for disambiguating information, or they may become more likely to recognize that social information can be a source of disambiguation.

We did not observe selective social referencing during the *label* phase, even when referential gaze was available. This result rules out the possibility that children are less selective during labeling because they learned that gaze direction was not informative. Instead, children may attend to someone who is speaking regardless of the need for disambiguation.

Social referencing and accuracy. In the metacognitive framework, confidence judgments are generally compared for correct and incorrect responses, with lower confidence for incorrect responses taken as evidence for metacognitive accuracy [Yeung & Summerfield (2012). Here, a parallel approach would be to examine the amount of social referencing children demonstrate for correct and incorrect responses.

To this end, we fit a mixed-effects linear regression model with the following structure: `number of looks ~ accuracy * age in months * phase + (1 | subject)`. The number of looks was collapsed across trial types and gaze conditions, since there were relatively few incorrect trials overall and none at all for some of the conditions. Results showed that accuracy and phase interacted such that children looked at the experimenter significantly more during the labeling phase of accurate trials (β = 0.4, $p < .001$), and significantly less in the *response* phase of accurate trials (β = -1.0, $p < .001$).

These findings confirm that children reference the experimenter more while responding incorrectly. Importantly, this analysis was restricted to trials in which there

was a correct response available (i.e., because the target was familiar or could be inferred by excluding the familiar distracter or observing the speaker's gaze). Thus, children's looking is sensitive not only to complete ignorance, but other factors that affect accuracy, which might include graded evidence or relative familiarity with the stimuli.

Additionally, children were more likely to answer correctly if they referenced the speaker during the labeling phase. This pattern could be due to children obtaining referential evidence from the experimenter's gaze in the gaze condition, or it could represent trials in which children were more engaged in the task.

General Discussion

During the preschool years, children are increasingly able to actively gather information through help-seeking and exploration (Chouinard et al., 2007; Schulz & Bonawitz, 2007). Do children monitor uncertainty in their knowledge to guide these behaviors, or are they indiscriminate with regard to underlying knowledge states? Here, we examined whether young children's social referencing during a word-learning task was driven by uncertainty about a label's referent.

We found that referential ambiguity strongly predicted children's social referencing. Specifically, we observed this selectivity when children were physically executing their decision, by placing the chosen object in a bucket in Experiments 1 and 2 and also while reaching for an object in Experiment 1. We speculate that children referenced the speaker during the decision process because they expected evaluative feedback about their demonstrated choice, either implicitly through the adult's facial expressions, or through an explicit response. This idea is consistent with other recent findings that preschoolers and toddlers seek help selectively when a problem is difficult or they are less skilled (Goupil et al., 2016; Vredenburgh & Kushnir, 2015).

Intriguingly, we also found that in the aggregate children's looking reflected graded referential evidence. In the case of FN trials, children could solve the problem of reference by excluding the familiar item (Markman & Wachtel, 1988; Merriman, Bowman, & MacWhinney, 1989), and indeed, they chose the correct object most of the time for these trials. If children simply monitored the presence or absence of such cues, they would have consistently responded to FN trials with certainty. Instead, their social referencing depended on a combination of cues, suggesting that preschoolers are sensitive to graded referential evidence. Meanwhile, children demonstrated uncertainty on trials with only one cue to reference (i.e., FN trials with no referential gaze and NN trials with referential gaze), suggesting that they remain subjectively uncertain about new label-object mappings if they have not received confirmation of its accuracy. These findings are important given that being able to monitor the likelihood of accuracy based on graded evidence is assumed important for decision-making and behavioral regulation across development (Krebs & Roebers, 2010; Yeung & Summerfield, 2012). For example, being able to monitor graded epistemic uncertainty allows individuals to gate out information that does not meet a criterion level of certainty based on individual goals or task demands (Koriat & Goldsmith, 1996).

These findings also bear on probabilistic models of learning and cognition, which predict that the proportion of individuals endorsing a belief will be proportional to the posterior probability of that belief based on Bayesian inference (e.g., E. Bonawitz et al., 2014; Denison, Bonawitz, Gopnik, & Griffiths, 2013; E. Vul, Goodman, Griffiths, & Tenenbaum, 2014). The current findings are consistent with the possibility that preschoolers might engage in this type of reasoning, given that their uncertainty tracked quantitatively with the amount of evidence available.

On the other hand, we found no evidence for selective social referencing while the

speaker was producing the label. One possibility is that young children do not recognize the need for disambiguating information until they need to make a decision (Markman, 1977). Another possibility is that preschool-aged children spontaneously look at a speaker regardless of ambiguity, and additional looking was not necessary or possible. This latter possibility seems more credible, given that children looked at the speaker at least once during labeling on most trials. Notably, Vaish et al. observed selective referencing during labeling among infants. Since infants in that study were holding one of the objects during labeling, referencing the speaker would have required them to disengage from that object, and may therefore have been more costly, promoting selectivity. Future research with preschoolers that includes a greater trade-off between attentional options would help to distinguish among these possibilities.

Overall, these results provide evidence that preschool-aged children monitor graded uncertainty in their mental representations. Furthermore, they act on that uncertainty through spontaneous information-seeking. These behaviors may underlie the rapid learning observed across domains in early childhood.

Acknowledgements

References

Balcomb, F. K., & Gerken, L. (2008). Three-year-old children can access their own memory to guide responses on a visual matching task. *Developmental Science*, *11*(5), 750–760.

Baldwin, D. (1991). Infants' contribution to the achievement of joint reference. *Child*

Development, *62*(5), 875–890.

Baldwin, D. (1993). Early referential understanding: infants' ability to recognize referential acts for what they are. *Developmental Psychology*, *29*, 832–843.

Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-Stay, Lose-Sample: A simple sequential algorithm for approximating Bayesian inference. *Cognitive Psychology*, *74*(C), 35–65.

Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition*, *3*(4), 207–220.

Chouinard, M. M., Harris, P. L., & Maratsos, M. P. (2007). Children's questions: A mechanism for cognitive development. *Monographs of the Society for Research in Child Development*, *72*, 1–129.

Corriveau, K., & Harris, P. L. (2009). Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information. *Developmental Science*, *12*(1), 188–193.

Coughlin, C., Hembacher, E., Lyons, K. E., & Ghetti, S. (2014). Introspection on uncertainty and judicious help-seeking during the preschool years. *Developmental Science*, *18*(6), 957–971.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of change in brain and cognitive development* (pp. 249–274). Oxford: Oxford University Press: Oxford University Press.

Daw, N. D., & Courville, A. C. (2007). The pigeon as particle filter. In *Advances in neural information processing systems*.

Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T. L. (2013). Rational variability in

children's causal inferences: The Sampling Hypothesis. *Cognition, 126*(2), 285–300.

Flavell, J. H., Green, F. L., Flavell, E. R., Harris, P. L., & Astington, J. W. (1995). Young children's knowledge about thinking. *Monographs of the Society for Research in Child Development*, i–iii–v–vi–1–113.

Ghetti, S., Hembacher, E., & Coughlin, C. (2013). Feeling Uncertain and Acting on It During the Preschool Years: A Metacognitive Approach. *Child Development Perspectives, 7*(3), 160–165.

Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they dont know. *Proceedings of the National Academy of Sciences, 113*(13), 3492–3496.

Gureckis, T. M., & Markant, D. B. (2012). Self-Directed Learning. *Perspectives on Psychological Science, 7*(5), 464–481.

Hembacher, E., & Ghetti, S. (2014). Dont look at my answer: Subjective uncertainty underlies preschoolers exclusion of their least accurate memories. *Psychological Science, 25*(9), 0956797614542273–1776.

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young Children Have a Specific, Highly Robust Bias to Trust Testimony. *Psychological Science, 21*(10), 1541–1547.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple Nor Too Complex. *PLoS ONE, 7*(5), e36399–8.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks Effect in Infant Auditory Attention. *Child Development, 19*, n/a–n/a.

Koenig, M. A., Clement, F., & Harris, P. L. (2009). Trust in testimony: Children's use

of true and false statements. *Psychological Science*, *15*, 694–698.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*(3), 490–517.

Krebs, S. S., & Roebers, C. M. (2010). Children's strategic regulation, metacognitive monitoring, and control processes during test taking. *British Journal of Educational Psychology*, *80*(3), 325–340.

Li, Q.-G., Heyman, G. D., Xu, F., & Lee, K. (2014). Young children's use of honesty as a basis for selective trust. *Journal of Experimental Child Psychology*, *117*(C), 59–72.

Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development*, *82*(6), 1778–1787.

Lyons, K. E., & Zelazo, P. D. (2011). Monitoring, metacognition, and executive function: Elucidating the role of self-reflection in the development of self-regulation. *Advances in Child Development and Behavior*, *40*, 379–412.

Markman, E. M. (1977). Realizing that you don't understand: A preliminary investigation. *Child Development*, *48*(3), 986–992.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121–157.

Merriman, W. E., Bowman, L. L., & MacWhinney, B. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, *54*, i–iii–v–1–129.

Ruggeri, A., Sim, Z. L., & Xu, F. (2017). "Why is Toma late to school again?" Preschoolers identify the most informative questions. *Developmental Science*.

Sabbagh, M. A., & Baldwin, D. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers theory of mind and semantic

development. *Child Development*, *72*, 1054–1070.

Schneider, W. (1998). Performance prediction in young children: Effects of skill, metacognition and wishful thinking. *Developmental Science*, 291–297.

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, *43*(4), 1045–1050.

Sim, Z. L., & Xu, F. (2017). Learning higher-order generalizations through free play: Evidence from two- and three-year-old children. *Developmental Psychology*.

Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review*, *120*, 779=797.

Sodian, B., Thoermer, C., Kristen, S., & Perst, H. (2012). Metacognition in infants and young children. In M. J. Beran, J. Brandl, J. Perner, & J. Proust (Eds.), *Foundations of metacognition* (pp. 119–133).

Sorce, J. F., Emde, R. N., Campos, J., & Klinnert, M. D. (1985). Maternal emotional signaling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology*, *21*, 195–200.

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants learning and exploration. *Science*, *348*(6230), 91–94.

Striano, T., & Rochat, P. (2000). Emergence of selective social referencing in infancy. *Infancy*, *1*, 253–264.

Tamis LeMonda, C. S., Adolph, K. E., Lobo, S. A., Karasik, L. B., Ishak, S., & Dimitropoulou, K. A. (2008). When infants take mothers' advice: 18-month-olds integrate perceptual and social information to guide motor action. *Developmental*

Psychology, *44*(3), 734–746.

Teglas, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure Reasoning in 12-Month-Old Infants as Probabilistic Inference. *Science*, *332*(6033), 1054–1059.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*, 1454–1463.

Vaish, A., Demir, ÃŰ. E., & Baldwin, D. (2011). Thirteen- and 18-month-old infants recognize when they need referential information. *Social Development*, *20*(3), 431–449.

Vredenburgh, C., & Kushnir, T. (2015). Young children's help-seeking as active information gathering. *Cognitive Science*, *40*(3), 697–722.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and Done? Optimal Decisions From Very Few Samples. *Cognitive Science*, *38*(4), 599–637.

Walden, T. A., & Ogan, T. A. (1988). The development of social referencing. *Child Development*, *59*(5), 1230–1240.

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1310–1321.

Zarbatany, L., & Lamb, M. E. (1985). Social referencing as a function of information source: Mothers versus strangers. *Infant Behavior and Development*, *8*, 25–33.

Predictor	Estimate	Std. Error	t value	p value	
Intercept	0.11	0.09	1.21	0.23	
Acc(Y)	0.01	0.09	0.06	0.95	
Age	-0.00	0.01	-0.30	0.76	
Phase(Label)	0.96	0.12	8.00	$< .001$	***
Phase(Planning)	0.10	0.12	0.81	0.42	
Phase(Response)	1.31	0.12	10.91	$< .001$	***
Acc(Y) * Age	0.00	0.01	0.21	0.84	
Acc(Y) * Phase(Label)	0.41	0.12	3.31	$< .001$	***
Acc(Y) * Phase(Planning)	-0.17	0.12	-1.33	0.18	
Acc(Y) * Phase(Response)	-0.99	0.12	-7.99	$< .001$	***
Age * Phase(Label)	0.04	0.02	1.90	0.06	.
Age * Phase(Planning)	0.02	0.02	0.99	0.32	
Age * Phase(Response)	0.02	0.02	0.98	0.33	
Acc(Y) * Age * Phase(Label)	-0.02	0.02	-1.16	0.25	
Acc(Y) * Age * Phase(Planning)	-0.02	0.02	-1.00	0.32	
Acc(Y) * Age * Phase(Response)	-0.03	0.02	-1.25	0.21	

Table 6

*Predictor estimates with standard errors and significance information for a linear
mixed-effects model predicting social referencing based on accuracy in Experiment 2.*

| Predictor	Estimate	Std. Error	t value	p value	
Intercept	0.14	0.06	2.62	0.01	**
Trial type(FN)	-0.09	0.07	-1.27	0.21	
Trial type(NN)	-0.02	0.08	-0.29	0.77	
Age	0.00	0.01	0.27	0.79	
Gaze(Y)	-0.01	0.08	-0.09	0.93	
Phase(Label)	1.36	0.07	18.71	< .001	***
Phase(Planning)	-0.11	0.07	-1.54	0.12	
Phase(Response)	0.10	0.07	1.38	0.17	
Trial type(FN) * Age	-0.01	0.01	-0.80	0.43	
Trial type(NN) * Age	-0.01	0.01	-0.66	0.51	
Trial type(FN) * Gaze	0.08	0.10	0.77	0.44	
Trial type(NN) * Gaze	-0.01	0.11	-0.10	0.92	
Age * Gaze	-0.00	0.01	-0.35	0.72	
Trial type(FN) * Phase(Label)	0.02	0.10	0.24	0.81	
Trial type(NN) * Phase(Label)	-0.02	0.10	-0.23	0.82	
Trial type(FN) * Phase(Planning)	0.21	0.10	2.01	0.04	*
Trial type(NN) * Phase(Planning)	0.28	0.10	2.77	0.01	**
Trial type(FN) * Phase(Response)	0.53	0.10	5.09	< .001	***
Trial type(NN) * Phase(Response)	0.68	0.10	6.52	< .001	***
Age * Phase(Label)	0.02	0.01	2.07	0.04	*
Age * Phase(Planning)	-0.00	0.01	-0.31	0.76	
Age * Phase(Response)	-0.00	0.01	-0.09	0.93	
Gaze * Phase(Label)	-0.06	0.10	-0.59	0.56	
Gaze * Phase(Planning)	-0.00	0.10	-0.00	1	
Gaze * Phase(Response)	0.08	0.10	0.79	0.43	
Trial type(FN) * Age * Gaze	0.01	0.02	0.46	0.65	
Trial type(NN) * Age * Gaze	0.01	0.02	0.49	0.62	
Trial type(FN) * Age * Phase(Label)	0.01	0.02	0.73	0.47	
Trial type(NN) * Age * Phase(Label)	-0.00	0.02	-0.19	0.85	
Trial type(FN) * Age * Phase(Planning)	0.00	0.02	0.04	0.96	
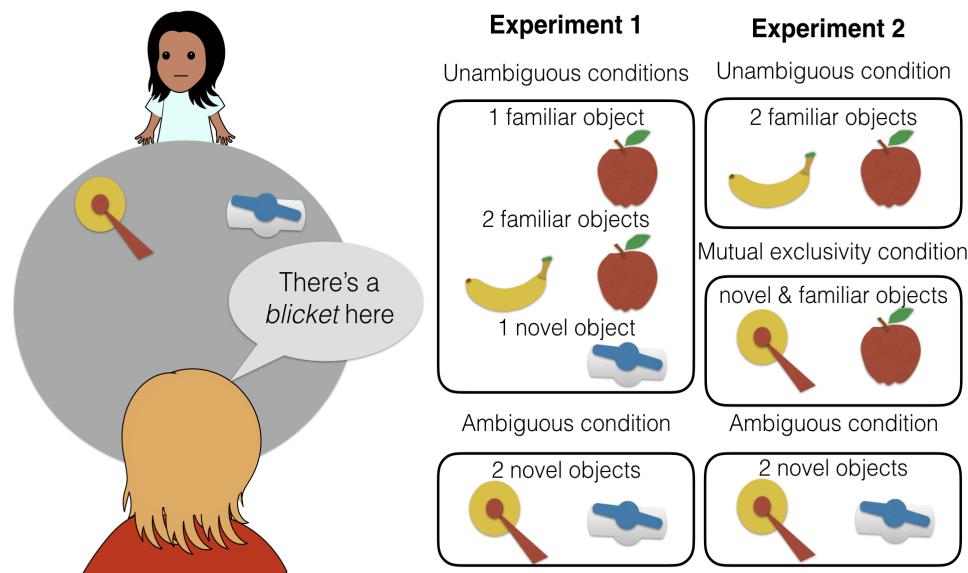
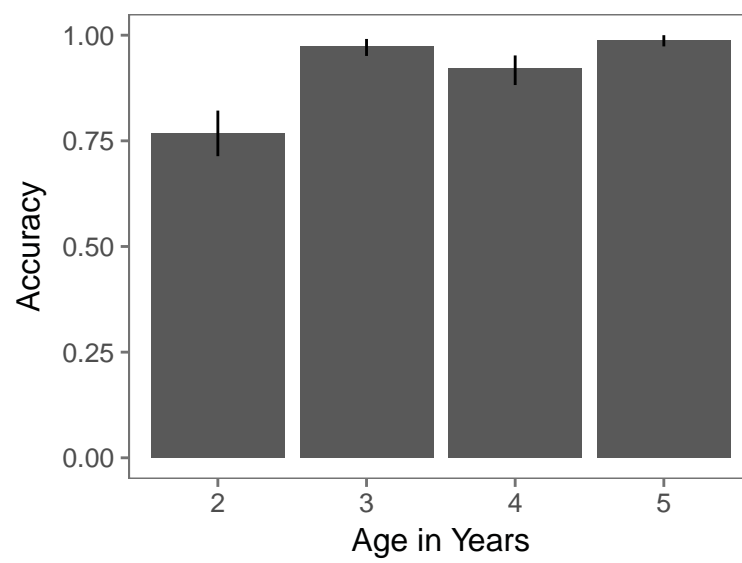*Figure 1*. Study design for experiments 1 and 2.

Figure 2. Accuracy for FF trials in Experiment 1. Error bars are 95 percent confidence intervals.
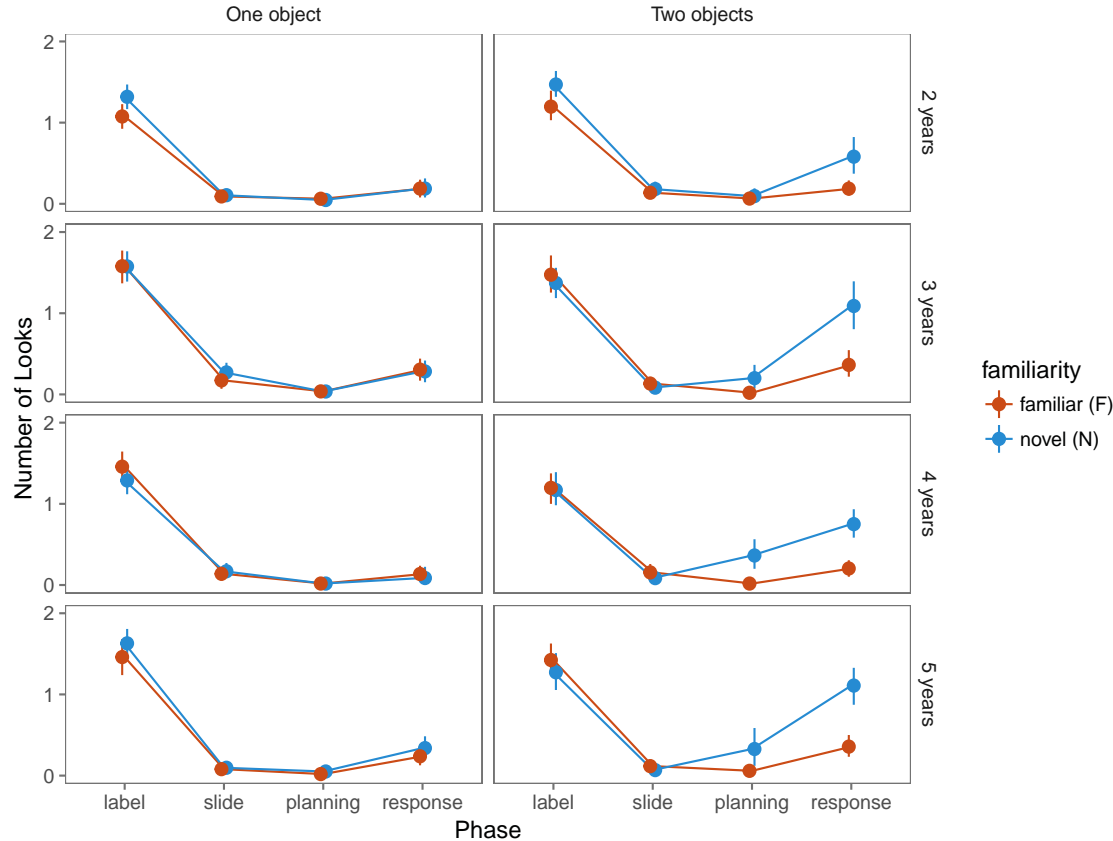
Figure 3. Results of Experiment 1. Number of looks to the experimenter based on phase, the number and familiarity of objects present, and age. Age in months was entered as a continuous variable in regression models but is presented here as a categorical variable for visual simplicity. Error bars are 95 percent confidence intervals.
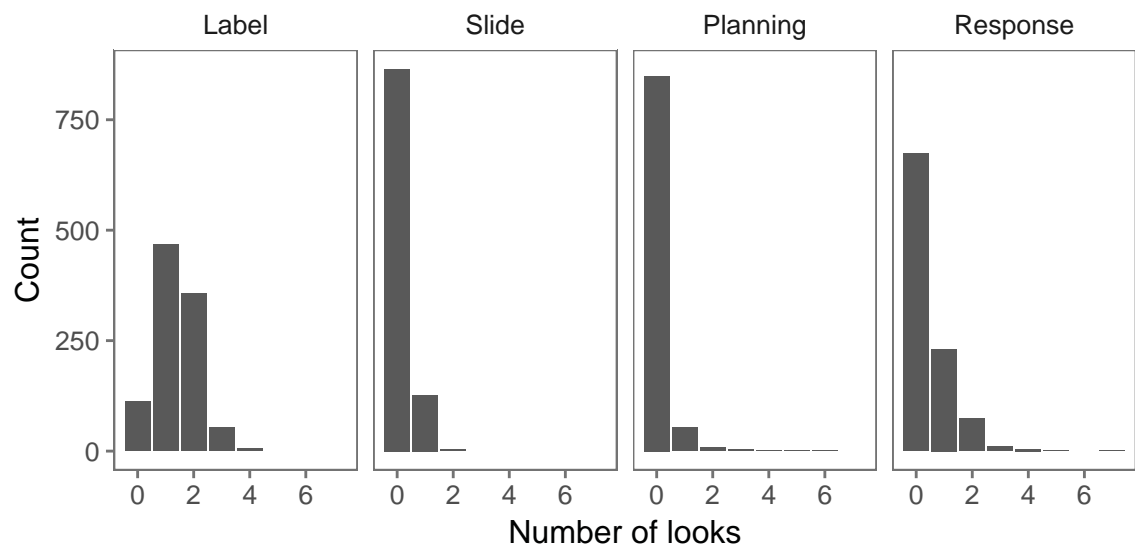
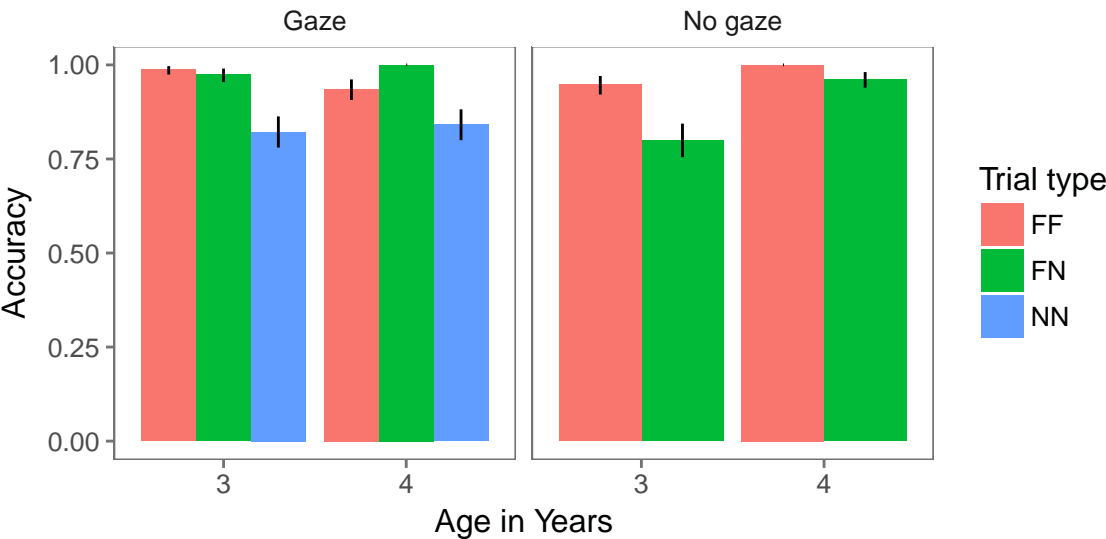*Figure 4.* Distribution of the number of looks to the speaker in each phase.

Figure 5. Accuracy for trials with a correct answer available (FF and FN, and all gaze trials) in Experiment 2. Error bars are 95 percent confidence intervals.
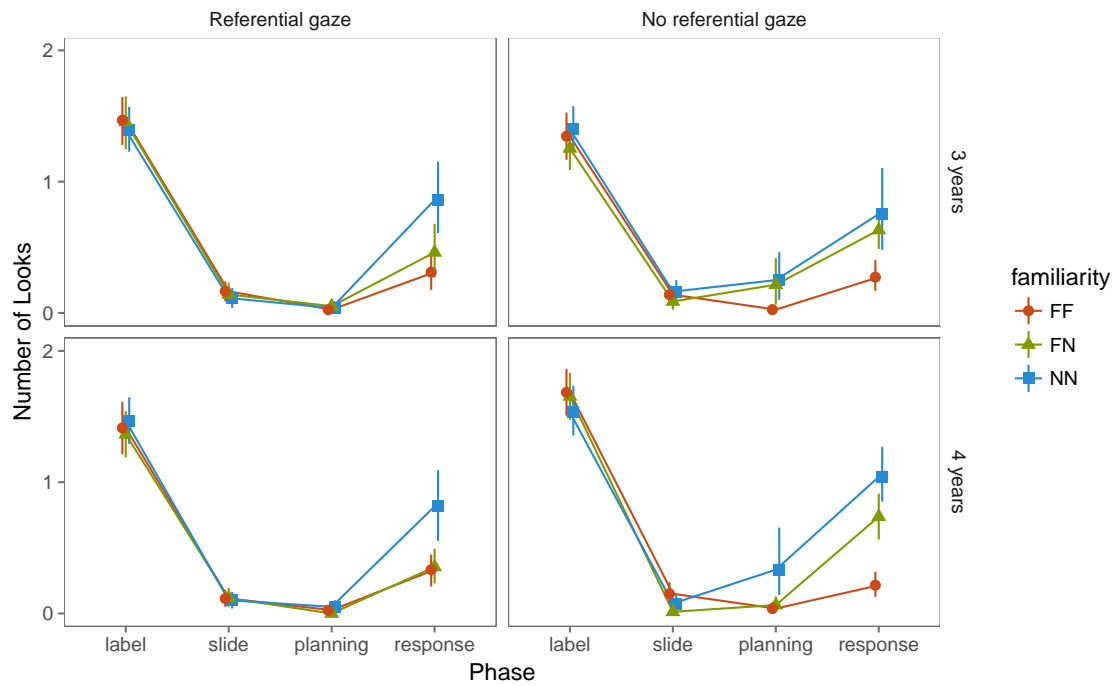
Figure 6. Results of Experiment 2. Number of looks to the experimenter based on phase, trial type, gaze condition, and age. Age in months was entered as a continuous variable in regression models but is presented here as a categorical variable for visual simplicity. Error bars are 95 percent confidence intervals.
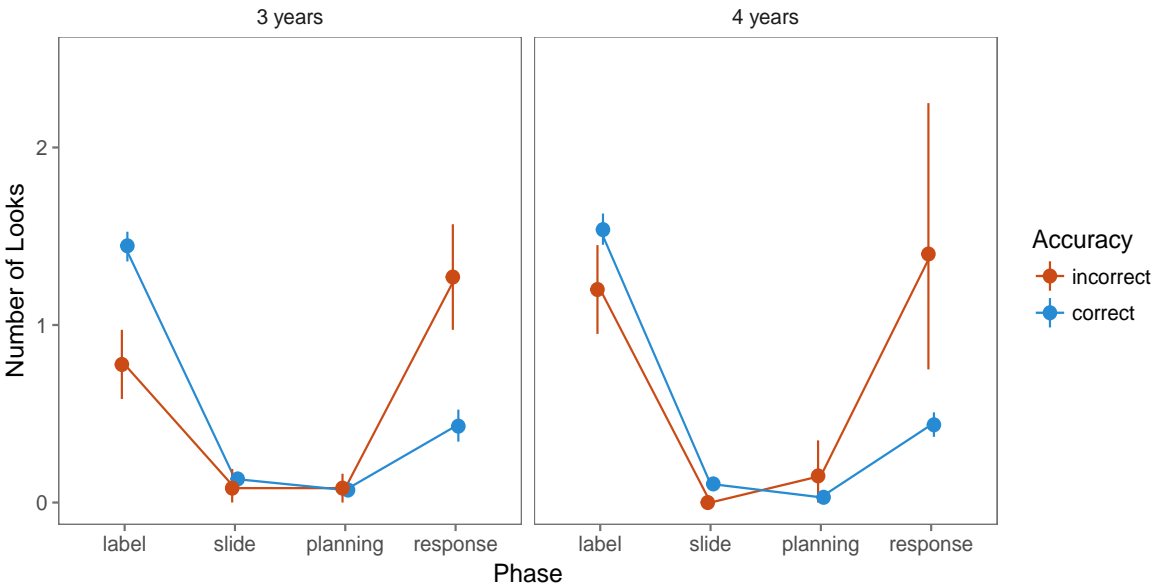
Figure 7. Results of Experiment 2. Number of looks to the experimenter based on accuracy, phase, and age, collapsed across trial types and gaze conditions. Age in months was entered as a continuous variable in regression models but is presented here as a categorical variable for visual simplicity. Error bars are 95 percent confidence intervals.